

K öln **u** nd **B** onner **A** rchaeologica

KuBA 1 / 2011

Kölner und Bonner Archaeologica

KuBA 1 / 2011

Herausgeber

Martin Bentz – Dietrich Boschung – Thomas Fischer –
Reinhard Förtsch – Michael Heinzelmann – Frank Rumscheid

Redaktion, Satz und Gestaltung

Torsten Zimmer, Jan Marius Müller und Stefanie Ostendorf

Umschlaggestaltung

Torsten Zimmer

Fotonachweis Umschlag

Gisela Geng (CoDArchLab), FA-SPerg000055-02

Alle Rechte sind dem Archäologischen Institut der Universität zu Köln und der
Abteilung für Klassische Archäologie der Rheinischen Friedrich-Wilhelms-Universität Bonn
vorbehalten. Wiedergaben nur mit ausdrücklicher Genehmigung.

Projekt: Image-Grid – Information und Bild. Automatisierte Kontexte

SEBASTIAN CUY – PAUL SCHEDING

Ziel des Projektes Image-Grid ist die automatisierte Verknüpfung von Bild und zugehöriger Information. Bei der Bearbeitung enormer Datenmengen – wie etwa den Bildbeständen der Fotothek des DAI Rom mit etwa 400.000 Bildern – ist eine Verknüpfung von Bildern mit wissenschaftlichen Daten per Hand so gut wie unmöglich. Innerhalb des Projekts Image-Grid werden anhand mehrerer Teilprojekte Wege entwickelt, die eine automatische Verbindung von Informationen ('Metadaten') mit den zugehörigen Bildern ermöglichen:

Im Zuge der Digitalisierung archäologischer Publikationen des frühen 20. Jhs. werden die erstellten Daten nicht als digitale Bücher betrachtet, sondern im Sinne des Projekts als »Lieferanten« wichtiger Information zu archäologischen Objekten. So werden anhand der Kataloge von Walter Amelung zu den Vatikanischen Museen² exemplarisch digitale Texte mit digitalen Fotografien der beschriebenen Objekte kontextualisiert. Über ein Texterkennungsverfahren, dem Optical Character Recognition (OCR), werden die Fotografien der Katalogseiten als Textdatei ausgelesen und dadurch nach bestimmten Begriffen, wie etwa Inventarnummern, durchsuchbar. Durch ein speziell modifiziertes 'Text-Mining'³ Programm können nun die Begriffe vom Text extrahiert werden; nach diesen Begriffen werden bereits bestehende Datensätze in der Datenbank ARACHNE durchsucht und bei Übereinstimmung der Inventarnummer miteinander verbunden. Den einzelnen Sei-

ten werden so Datensätze und damit Bilder in der ARACHNE zugeordnet, die exakt auf dieser digitalen Buchseite beschrieben sind.

Zum zweiten Projekt gehört die digitale Weiterverarbeitung der Inventarbücher der Fotothek des DAI Rom, die seit Bestehen der Fotothek 1928 geführt werden. Zur Sicherung der einmaligen Dokumente wurden diese digitalisiert und über den Bookbrowser⁴ der ARACHNE-Datenbank veröffentlicht. In den etwa 100 Büchern sind Namen und Beschreibungen zu den jeweiligen Negativen verzeichnet, die in den Jahren 1928 bis 1999 erstellt wurden. Jede Negativnummer ist einmalig, um die Bilder jederzeit identifizieren zu können. Da diese Negativnummer auch bei der Digitalisierung der analogen Bestände des DAI Rom als Name der Scans verwendet werden (siehe Projekt EMAGINES), ist es möglich die Bilder automatisiert mit den Seiten des Inventarbuches und den darauf vermerkten Informationen zu verlinken. Zusätzlich können über diese Negativnummer bereits digitalisierte archäologische Publikationen, wie etwa durch Google Bücher⁵ durchsucht und automatisch mit dem Bild verbunden werden.

Zu Beginn der 1990er Jahre wurden in Rom die gesamten Bildbestände der Fotothek (ca. 250.000 Bilder) als Mikroplanfilm ('Microfiche') archiviert⁶. Um die Bilder auf den Microfiche-Filmen zu identifizieren wurde ein Indexband erstellt, der als Konkordanz zwischen Microfiche und realer Struktur der Fotothek in Rom fungierte. Innerhalb

1 Der Text basiert auf dem Vortrag zum Projekt im Rahmen eines Kolloquiums: Tagungsbericht Depot und Plattform. Bildarchive im post-fotografischen Zeitalter. 05.06.2009–07.06.2009, Köln, in: H-Soz-u-Kult, zuletzt aktualisiert am 17.08.2009 <<http://hsozkult.geschichte.hu-berlin.de/tagungsberichte/id=2746>> (20.05.2010).

2 W. Amelung, Die Sculpturen des Vaticanischen Museums im Auftrage und unter Mitwirkung des Kaiserlich Deutschen Archäologischen Instituts (Römische Abteilung) Bd. I–III (Berlin 1903–1956).

3 Zum Verfahren vgl. den Eintrag Text-Mining, zuletzt aktualisiert am 27.04.2010 <http://de.wikipedia.org/Text_Mining> (20.05.2010).

4 iDAI.bookbrowser <<http://www.arachne.uni-koeln.de/drupal/?q=de/books>> (20.05.2010).

5 Google Bücher <<http://books.google.de>> (20.05.2010).

6 Die eigentliche Systematik, in der die Fotografien in der Fotothek Rom strukturiert sind, konnte dabei auf Grund der Form der Microfiche in Format DIN A6 nicht beibehalten werden, da maximal 98 Bilder auf einem Fiche gespeichert werden können.

des Projekt Image-Grid wurden die Microfiche automatisch hochauflösend gescannt und jede Fotografie auf dem Fiche als einzelnes Bild gespeichert. Die Seiten des Indexbands wurden wiederum über OCR als Text-Datei ausgelesen. Anhand des Textes war es über ein eigenes hierfür entwickeltes Programm möglich, die einzelnen Bilder der Fotothek wieder in ihre eigentliche Struktur zurückzuführen und deren hierarchische Systematik wiederherzustellen. Die Struktur der Fotothek ist nun über einen neu entwickelten Browser durchsuchbar, der bis Ende des Jahres 2010 der Öffentlichkeit prä-

sentiert werden soll. Die einzelnen Projekte fließen schlussendlich in der Datenbank ARACHNE zusammen, indem sowohl die Inventarbücher des DAI als auch die digitalen Publikationen mit den einzelnen Bildern automatisch verbunden werden. Der Nutzer kann so vom Bild bzw. archäologischen Objekt auf die Informationen aus ARACHNE, digitalen Büchern und Inventarbüchern direkt zugreifen. Damit wurde ein wichtiger Schritt in Richtung einer integrierten, digitalen Forschungsumgebung getan, die textuelle und fotografische Quellen der Archäologie verbindet.